

AD-A113 027

PITTSBURGH UNIV PA DEPT OF MATHEMATICS AND STATISTICS  
MULTIVARIATE ANALYSIS AND ITS APPLICATIONS. (U)

F76 12/1

DEC 81 P R KRISHNAIAH

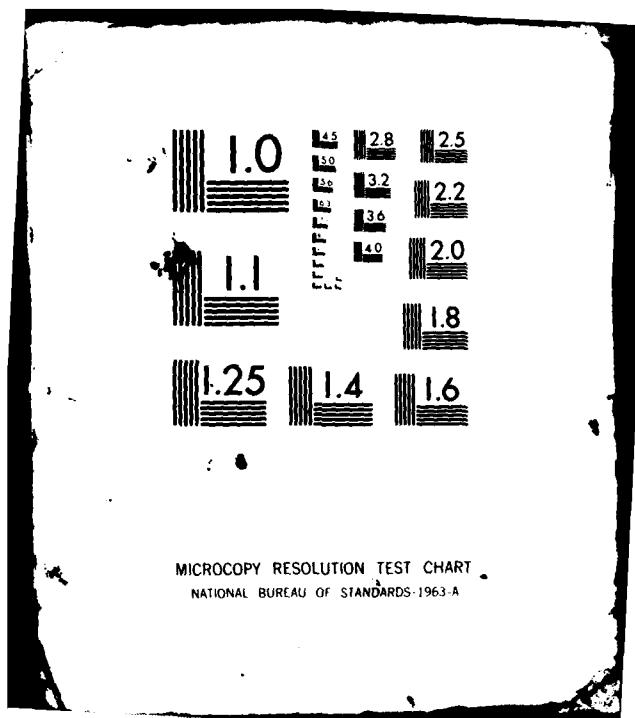
F49620-79-C-0161

UNCLASSIFIED

AFOSR-TR-82-0271

NL

END  
DATE  
FILED  
4 82  
DTIG



AFOSR-TR- 82-0271

②

AD A11 3027

MULTIVARIATE ANALYSIS AND ITS  
APPLICATIONS

FINAL REPORT

Contract F49620-79-C-0161  
June 1, 1979 - Dec. 14, 1981

DTIC FILE COPY

DTIC  
ELECTED  
APR 6 1982  
S D

Approved for public release;  
distribution unlimited.

82 04 06 068

*7/10/78*  
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER <b>APOSR-TR- 82-0271</b>	2. GOVT ACCESSION NO. <b>AD 4113027</b>	3. DOCUMENT'S CATALOG NUMBER
4. TITLE (and Subtitle) <b>MULTIVARIATE ANALYSIS AND ITS APPLICATIONS</b>	5. TYPE OF REPORT & PERIOD COVERED <b>Final</b>	
6. AUTHOR(s) <b>Prof. Krishnaiah</b>	7. CONTRACT OR GRANT NUMBER <b>F49620-79-C-0161</b>	
8. PERFORMING ORGANIZATION NAME AND ADDRESS <b>University of Pittsburgh Dept. of Math &amp; Statistics Pittsburgh, PA 15260</b>	9. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS <b>61102F 2304/A3</b>	
10. CONTROLLING OFFICE NAME AND ADDRESS <b>APOSR/NM Bolling AFB, DC 20332</b>	11. REPORT DATE <b>6/18/81</b>	
12. MONITORING AGENCY NAME & ADDRESS if different from Controlling Office	13. NUMBER OF PAGES <b>24</b>	
14. SECURITY CLASS (or this report) <b>Unclassified</b>		
15. FUNDING AGENCY SOURCE(S) <b>DOE</b>		
16. DISTRIBUTION STATEMENT (for this Report) <b>Approved for public release; distribution unlimited</b>		
17. DISTRIBUTION STATEMENT (for the abstract entered in Block 18, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) <p>This report covers the work done under the contract F49620-79-C-0161 during the period of June 1, 1979 - Dec. 14, 1981. All this work is reported in various papers and the abstracts of these papers are attached. The work in these papers is supported completely or partially by the above contract.</p>		

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 69 IS OBSOLETE

*7/10/78*  
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

This report covers the work done under the contract  
F49620-79-C-0161 during the period of June 1, 1979 -  
Dec. 14, 1981. All this work is reported in various papers  
and the abstracts of these papers are attached. The work  
in these papers is supported completely or partially by the  
above contract.

Assessment Per	
NSRIS Status	X
NSRIS ZDR	O
Unpublished	
Justification	
By _____	
Distribution/	
Availability Codes	
Small and/or	
Large	Special

070  
DOD  
DTIC

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFOSR)  
NOTICE OF SUBMISSION TO DTIC  
This technical report has been reviewed and is  
approved for public release AF/AFS 100-10.  
Distribution is unlimited.  
ROBERT J. KERSEY  
Chief, Technical Information Division

(11) Block, H. W. and Thompson, A. B. (1982). Inequalities on distributions: bivariate and multivariate. To appear in Encyclopedia of Statistical Sciences (N. L. Johnson and S. Kotz, editors). John Wiley and Sons.

Inequalities concerning bivariate and multivariate distributions in statistics are surveyed, as well as historical background. Subjects treated include inequalities arising through positive and negative dependence, Boole, Bonferroni and Prechet inequalities, convex symmetric set inequalities; stochastic ordering; stochastic approximation and inequalities obtained by majorization; Chebychev and Kolmogorov type inequalities; multivariate moment inequalities; and applications to simultaneous inference, unbiased testing and reliability theory.

(12) Dubois, J. and Rao, C. R. (1980). On the convexity of some divergence measures based on entropy functions. IEEE Information Theory, 1981, to press.

Three measures of divergence between vectors in a convex set of  $n$ -dimensional real vector space have been defined in terms of certain types of entropy functions, and their convexity property studied. Among other results, a classification of the entropies of degree  $\alpha$  is obtained by the convexity of these measures. These results have applications in information theory and biological studies.

(13) Dubois, J. and Rao, C. R. (1980). Entropy differential metric, distance and divergence measures in probability spaces - a unified approach. Technical Report No. 80-10, Institute for Statistics and Applications, Department of Mathematics and Statistics, University of Pittsburgh.

The paper is devoted to unification of probability spaces through the introduction of a quadratic differential metric in the parameter space of the probability distributions. For this purpose, a  $\theta$ -entropy functional is defined on the probability space and its gradient along a direction of the tangent space of the parameter space is taken as the metric. The distance between two probability distributions is computed as the geodesic

distance induced by the metric. The paper also deals with three measures of divergence between probability distributions and their inter-relationships.

16) Burbea, J. and Rao, C. R. (1981). On the convexity of higher order Jensen differences based on entropy functions. Technical Report No. 81-20, Institute for Statistics and Applications, Department of Mathematics and Statistics, University of Pittsburgh.

In an earlier work, the authors introduced a divergence measure, called the first order Jensen difference, or the  $\beta$ -divergence, which is based on entropy functions of degree  $\alpha$ . This provided a generalization of the measure of output information based on Shannon's entropy (corresponding to  $\alpha = 1$ ). It was shown that the  $\beta$ -order  $\beta$ -divergence is a convex function with respect to  $\beta$  if  $\beta$  belongs to some range. In this paper we define higher order Jensen differences and show that they are convex functions only when the underlying entropy function is of degree  $\beta$ . A statistical application regarding the convexity of higher order Jensen differences is presented.

17) Burbea, J. and Rao, C. R. (1981). Differential metric in statistical spaces. Technical Report No. 81-29, Institute for Statistics and Applications, Department of Mathematics and Statistics, University of Pittsburgh.

We discuss the construction of differential metrics in statistical spaces through entropy functions and examine their relationship with the Information metric introduced by Rao using the Fisher Information metric in the statistical studies of classification and discrimination, and the classical Bergman metric. It is suggested that the scalar and field curvatures associated with the Bergman-Information metric may yield results in statistical inference analogous to those of Riemann using the Gaussian curvature.

16) Chibase, T. (1980). Invariant polynomials with real and complex matrix arguments and their applications. Technical Report No. 80-3. Institute for Statistics and Applications, Department of Mathematics and Statistics, University of Pittsburgh.

A class of homogeneous polynomials  $C_{ij}^{(r)}$ ,  
 $(i_1, i_2, \dots, i_r)$  with a general number  $r$  complex matrix arguments ( $i_j \in \mathbb{C}^n$ ) invariant under the orthogonal group, is proposed through the theory of polynomial representations of the linear group. The class of polynomials is a generalization of the invariant polynomials with smaller number of matrix arguments developed previously up to the case of three matrix arguments. Properties, generation and relations of the polynomials  $C_{ij}^{(r)}$  are shown, and some applications in multivariate distribution theory are indicated. The complex analogue of the above statements is also illustrated.

17) Chibase, T. (1980). Simultaneous normal confidence regions for ratios of the discriminant coefficients. Technical Report No. 80-4. Institute for Statistics and Applications, Department of Mathematics and Statistics, University of Pittsburgh.

We consider simultaneous confidence regions for some hypotheses on ratios of the discriminant coefficients of the linear discriminant function when the population mean and common covariance matrix are unknown. This problem, involving hypotheses on ratios, yields the so-called "pseudo" confidence regions valid conditionally to subspace of the parameter space. We state the explicit formulae of the regions and give further discussions on the validity of these regions. Illustrations of the pseudo confidence regions are given.

18) Chibase, T. (1980). Distributions of some matrix variates and latent roots in multivariate DeMoivre-Polya distributions. Technical Report No. 80-12. Institute for Statistics and Applications, Department of Mathematics and Statistics, University of Pittsburgh.

In this paper, it is shown that some distributions

of the inverse variances and losses come arising in the multivariate Behrens-Fisher discrimination problem can be explicitly expressed in terms of the invariant polynomials over the matrix arguments, due to A. N. Dorroh, extending the usual polynomials of suffice arguments.

(10) CRAMER, R. (1946). Substitution of a test statistic under the Gauss-Markov model. Technical Report No. 00-14, Institute for Statistics and Applications, Department of Mathematics and Statistics, University of Pittsburgh.

A statistical procedure based on the simple Gauss-Markov model ( $\mathbf{A}, \mathbf{B}, \sigma^2$ ) is said to be  $\alpha$ -imprecise, if it is uniformly worse under the general model ( $\mathbf{A}, \mathbf{B}, \sigma^2$ ). This substitution statistic has been considered in an alternative procedure combining the GMM and simple testing procedures in the form of the Gauss-Markov model. In this paper, we shall investigate the substitution of the likelihood ratio test statistic by the test of hypotheses on sufficient linear functions and also a modification of it for  $t_1 = t_2$ . This is a generalization of the result due to Ghosh and Bhattacharya (1971). We established the general case when  $\mathbf{A}$  has an interaction covariance structure. We also consider the substitution for a class of quadratic unbiased estimator equations as a special case.

(11) CRAMER, R. (1949). Representations of the covariance matrix for sufficiency in the Gauss-Markov model. Technical Report No. 00-13, Institute for Statistics and Applications, Department of Mathematics and Statistics, University of Pittsburgh.

This paper establishes some extensions of the results of the 1946 and the 1949 (1949 and 1951, Chapter 6), that give a table of general representations of the covariance matrix in terms of the given design matrix, under which various statistical procedures in the form of the theory based on the simple Gauss-Markov model and the orthogonal covariance matrix are also valid under the general Gauss-Markov. We shall also extend tables giving some more results relating to sufficiency, especially in connection with the estimation and testing of hypotheses on linear parametric functions.

(11) Chatterjee, S. (1980). Pseudo confidence regions for the estimation of a multivariate linear functional relationship. Technical Report No. 80-16. Institute for Statistics and Applications, Mathematics and Statistics Department, University of Pittsburgh.

The problem of finding confidence regions (CR) for a regression vector  $\beta$  given no the existence of a linear functional relationship (LFR) is not investigated. Note an ordered vector  $\beta$  and an ordered set  $\{x_i\}_{i=1}^n$ ,  $x_i$  is an  $m$ -dimensional random variable and  $x_i \neq x_j$  for  $i \neq j$ . Then  $\beta' = (1, x_1', \dots, x_n')$  is natural.  $\beta'$  is called the  $\beta$ -vector and  $x_i$  is called the  $i$ -th  $x$ -vector.  $\beta'$  is called the  $\beta$ -vector and  $x_i$  is called the  $i$ -th  $x$ -vector. Note that  $\beta$  is a generalization of the unknown function used for the type of output given  $y = \beta' x + \epsilon$  at level of confidence  $1 - \alpha$ . To obtain a confidence interval, which shall be the so-called "pseudo" confidence region (PCR) called confidence interval of the  $\beta$ -vector where the distribution is based on the "pivotal vector" confidence factor (PCF) about by the method of a pivotal statistic. The PCR conditions necessary for a PCR to exist are shown to be the consistency conditions concerning the sufficient LFR. The probability that these conditions hold approaches one under "reasonable circumstances" by their different authors. we can have a PCR with confidence coefficients  $1 - \alpha$ . Some simulation results are present.

(12) Hogg, R. V. and Tanis, J. I. (1983). On a general condition for test sufficiency. Accepted for publication in the Journal of Multivariate Analysis subject to revision.

This paper gives sufficient conditions that certain statistics have a common distribution under a wide class of underlying distributions. Sufficiency methods are the primary statistical tools used in establishing the theoretical results. These results are applied to  $\chi^2$ -type statistics, statistics involving corrected correlations, and certain statistics associated with the complex normal distribution.

[13] Fong, S. C. (1981). A survey of approximate distributions of functions of the eigenvalues of random matrices for nonnormal populations. Invited paper presented at the Eastern Regional Meeting of the Institute of Mathematical Statistics held in Philadelphia, May 1981.

In this paper, the author gave a review of the literature on various developments on certain aspects of multivariate distribution theory when the underlying distribution law is not multivariate normal.

[14] Fong, S. C. and Guttmann, R. A. (1981). Approximate distributions of functions of the eigenvalues of the doubly noncentral F matrix and their applications. Submitted for publication.

Let  $\lambda_1$  and  $\lambda_2$  be distributed independently as non-central Student  $t$  variables. Then the distribution of  $\lambda_1 \lambda_2^{-1}$  is known to be double noncentral F matrix. In this paper, approximate distributions of various functions of the eigenvalues of  $\lambda_1 \lambda_2^{-1}$  are derived. Applications of these results in the area of distributions of tests and other areas are discussed.

[15] Fong, S. C., Guttmann, R. A. and Guttmann, D. G. (1981). Distributional distributions of the likelihood ratio test statistic for nonhomogeneous estimation of the complex multivariate normal distribution. Accepted for publication, subject to revision in Journal of Multivariate Analysis.

In this paper, the authors obtained exact and approximate for the null and nonnull distributions of the likelihood ratio test statistic for multivariate independence when the underlying distribution is complex multivariate normal. The authors also derived approximate null distribution of the likelihood ratio test statistic for homogeneous of the covariance matrices of several related multivariate normal populations; approximate distributions are also obtained for the nonnull case when we have two populations. The distributions obtained in this paper are in terms of beta statistic.

I.6.1 Fang, C., Krishnamoorthy, P. R. and Rosenblatt, S. B. (1981). Asymptotic distribution of the likelihood ratio test statistic for hypotheses of complex multivariate normal distribution. Technical Report No. 81-77. Institute for Statistics and Applications, Mathematics and Statistics Department, University of Pittsburgh.

In this report, the authors investigated the distribution of certain power of the likelihood ratio test statistic for hypotheses of the complex multivariate normal distribution with a beta type series in the null and nonnull cases. Applications of these results to multivariate case are also discussed.

I.6.2 Fang, C. and Krishnamoorthy, P. R. (1982). Asymptotic distributions of functions of the eigenvalues of non random matrices for canonical correlations. Journal of Multivariate Analysis. (In press).

In this paper, the authors investigated the asymptotic joint distributions of certain functions of the eigenvalues of the sample covariance matrix, correlation matrix, and canonical correlation matrix in nonnull situations when the population eigenvalues have multisettles. These results are derived without assuming that the underlying distribution is multivariate normal. In obtaining these distributions, the authors used Edgeworth type distributions.

I.6.3 Fang, C. and Krishnamoorthy, P. R. (1983). Asymptotic distributions of functions of the eigenvalues of the test and complex noncentral Wishart matrices. In Multivariate and Related Inference (George, R., Dickey, G., Rao, J. S. S. and Sethuram, A. T. M. G., editors). Butterworth Publishing Company.

In this paper, the authors investigated asymptotic distributions of functions of the eigenvalues of the non-central Wishart matrix. Applications of these results are discussed to a studies in cluster analysis and to the problem of studying the structure of interaction in two-way classification with one observation per cell.

[14] Fang, C. and Krishnaiah, P. R. (1982). Asymptotic joint distributions of functions of the elements of sample covariance matrix. In Statistics and Probability: Essays in Honor of C. R. Rao. (G. Kallianpur, P. R. Krishnaiah and J. K. Ghosh, editors). North-Holland Publishing Company.

In this paper, the authors give asymptotic expressions for the joint distribution of the functions of the elements of the sample covariance matrix and sample correlation matrix in the noncentral case when the underlying distribution is multivariate normal. Accuracy of the above expressions is also studied. Also, asymptotic expressions are given for functions of the elements of the sample covariance matrix for nonnormal populations. Finally, some applications of the above results are discussed.

[15] Jonckheere, R. G. (1981). On characterizing the Rether-Polya distribution. Technical Report No. 81-00. Institute for Statistics and Applications, Department of Mathematics and Statistics, University of Pittsburgh.

A discrete survival model is considered where an unobservable random variable is subjected to destruction so that what is observed and recorded is only the undamaged part  $t$  of it. Assuming the destruction process is represented by the Rether-Polya distribution, a characterization of the negative binomial distribution is obtained. Utilizing the completeness property of the negative binomial distribution, a characterization of the Rether-Polya distribution is derived. Several other characterization theorems are also proved concerning these probability distributions.

[16] Jonckheere, R. G. and Rao, B. Rao (1981). Characterizations of generalized Rether-Polya and generalized Polya-Eggenberger distributions. Technical Report No. 81-03. Institute for Statistics and Applications, Department of Mathematics and Statistics, University of Pittsburgh.

A discrete model is considered where the original observation is subjected to partial destruction according to

the generalized Markov-Polya damage model. A characterization of the generalized Polya-Eggenberger distribution is given in the context of the Rao-Rubin condition. Several other characterization theorems are also proved concerning these probability distributions.

[12] Kariya, T. (1980). Robustness of multivariate tests. Annals of Statistics.

This paper gives necessary and sufficient condition for the null distribution of a test statistic to remain the same in the class of left  $O(n)$ -invariant distributions. Secondly, it is shown that in certain special cases, the usual MANOVA tests are still uniformly most powerful invariant in a class of left  $O(n)$ -invariant distributions.

[13] Kariya, T. (1981). A new concept of second order efficiency and its application to a missing data problem. Technical Report No. 81-16. Institute for Statistics and Applications, Department of Mathematics and Statistics, University of Pittsburgh.

In this paper, following the framework of Kariya, Krishnamoorthy and Rao (1981), a finite sample concept of second order efficiency is defined and it is applied to the analysis of the problem of estimating bivariate normal parameters with extra data on the first variate. Only the modified maximum likelihood estimator is shown to be second order efficient.

[14] Kariya, T., Krishnamoorthy, P. R. and Rao, C. R. (1981). Inference on the parameters of multivariate normal population when some data is missing. In Developments in Statistics, Vol. 4 (in press). Academic Press, New York.

In this paper, the authors considered the problems of estimation and tests of hypotheses on the parameters of the multivariate normal population when the data is incomplete. Special emphasis is made for the case of the bivariate normal when the variances of the variables are equal. The authors also obtained certain expressions for the asymptotic distributions of a wide class of test statistics useful in testing various hypotheses on the parameters of the multivariate normal when some of the data is missing.

[25] Kariya, T., Sinha, B. K. and Subramanyam, K. (1981).  
Nearly efficient estimators based on order statistics.  
Technical Report No. 81-05. Institute for Statistics  
and Applications, Department of Mathematics and Statistics,  
University of Pittsburgh.

In this paper, based on 3 or 5 order statistics out of  $n$  order statistics, nearly asymptotically efficient estimators are obtained for Cauchy, logistic and normal distributions.

[26] Kariya, T., Sinha, B. K. and Subramanyam, K. (1981).  
First, second and third order efficiencies of the estimators for a common mean. To be given for typing.

Based on Kariya, Krishnaiah and Rao (1981) and Kariya (1981), this paper considers efficiencies of several estimators proposed in the problem of estimating a common mean of  $K$  univariate normal populations. Only the Graybill-Deal (1959) estimator is shown to be third order efficient.

[27] Kariya, T., Sinha, B. K. and Krishnaiah, P. R. (1981).  
Some properties of left orthogonally invariant distributions. Technical Report No. 81-28. Institute for Statistics and Applications, Department of Mathematics and Statistics, University of Pittsburgh.

In this note, we have considered generalizations of some of the results on spherical distributions to the case of left spherical matrix variate distributions. First, it is shown that the independence of  $n$  rows of an  $n \times p$  left spherical random matrix implies multivariate normality. Secondly, most results in Eaton (1981) are extended to the matrix variate case.

[28] Kariya, T., Sinha, B. K. and Subramanyam, K. (1981).  
Berkson's bioassay problem - revisited. Technical Report No. 81-24. Institute for Statistics and Applications, Department of Mathematics and Statistics, University of Pittsburgh.

Consider Berkson's problem of estimating  $\theta$  on the basis of independent random variables  $X_i$  having the binomial distribution  $B(n_i, \pi_i(\theta))$  ( $i = 1, \dots, k$ ), where

$\pi_i(\theta) = [1 + \exp(-\theta - \beta d_i)]^{-1}$ , and  $\beta$  and  $d_i$ 's are known. Extensive asymptotic and nonasymptotic comparisons of two particular estimates of  $\theta$ , the MLE  $\hat{\theta}_M$  and the Rao-Blackwellized version  $\hat{\theta}_B$  of Berkson's estimator  $\hat{\theta}_B$ , have been made in the literature. Our subject in this paper is three fold: (a) to derive a simple necessary condition for  $\hat{\theta}_B$  to dominate  $\hat{\theta}_M$  uniformly in the MSE criterion; (b) to propose three new estimators and compare them with  $\hat{\theta}_B$  and  $\hat{\theta}_M$ ; (c) to show that  $\hat{\theta}_B$  and  $\hat{\theta}_M$  are the first order efficient no matter what  $\pi_i(\theta) = F(\theta + \beta d_i)$  may be so long as  $F$  is strictly increasing.

[29] Khatri, C. G. and Rao, C. R. (1981). Some extensions of the Kantorovich inequality and statistical applications. J. Multivariate Analysis, 11.

Kantorovich gave an upper bound to  $(x'Vx)(x'V^{-1}x)$  where  $x$  is an  $n$ -vector of unit length and  $V$  is an  $n \times n$  positive definite matrix. Bloomfield, Watson and Knott found the bound to  $|X'V X X'V^{-1}X|$ , and Khatri and Rao to the trace and determinant of  $X'V Y Y'V^{-1}X$  where  $X$  and  $Y$  are  $n \times k$  matrices such that  $X'X = Y'Y = I$ . In the present paper we establish bounds for traces and determinants of  $X'V Y Y'V^{-1}X$  and  $X'BYY'CX$  when  $X$  and  $Y$  are matrices of different orders. A review of previous results on generalizations of Kantorovich inequality and a number of new results of independent interest are also given.

[30] Kimeldorf, C., May, J. and Sampson, A. R. (1980). Concordant and discordant monotone correlations and their evaluation by nonlinear optimization. TIMS Studies in the Management Sciences: Optimization in Statistics Volume (to appear).

This paper presents four new statistical measures of monotone relationship derived from the concept of monotone correlation. A nonlinear optimization algorithm is employed to evaluate these new measures, as well as the monotone correlation, for ordinal contingency tables. A computer program to implement the algorithm is developed, and is applied to several insightful examples to provide further understanding of the usefulness of these measures.

[31] Kimeldorf, G., May, J. and Sampson, A. (1981). MONCOR - A program to compute concordant and other monotone correlations. Computer Science and Statistics: Proceedings of the 13th Symposium on Interface. Springer-Verlag, New York.

The new interactive FORTRAN program MONCOR is described. MONCOR computes the concordant monotone correlation, discordant monotone correlation, isoconcordant monotone correlation, isodiscordant monotone correlation and their associated monotone variables. Data input can be finite discrete bivariate probability mass functions or ordinal contingency tables, both of which must be given in matrix form. The well-known British Mobility data are used to illustrate the input and output options available in MONCOR.

[32] Kimeldorf, G., Plachky, D. and Sampson, A. R. (1981). A simultaneous characterization of the Poisson and Bernoulli distributions. J. Appl. Probability 18, 316-320.

Let  $N, X_1, X_2, \dots$  be non-constant independent random variables with  $X_1, X_2, \dots$  being identically distributed and  $N$  being nonnegative and integer valued. It is shown that the independence of  $\sum_{i=1}^N X_i$  and  $N - \sum_{i=1}^N X_i$  implies that the  $X_i$ 's have a Bernoulli distribution and  $N$  has a Poisson distribution. Other related characterization results are considered.

[33] Krishnaiah, P. R. (1981). Computations of some multivariate distributions. Handbook of Statistics, Vol. 1 (P. R. Krishnaiah, editor), 745-971. North-Holland Publishing Company.

This paper reviews the present state-of-the-art on computations of functions of the roots of several random matrices which arise in multivariate statistical analysis. Many of the available tables useful in statistical analysis of multivariate data are given in the Appendix.

[34] Krishnaiah, P. R. and Lee, J. C. (1981). Likelihood ratio tests for mean vectors and covariance matrices. Handbook of Statistics, Vol. 1. (P. R. Krishnaiah, editor). North-Holland Publishing Company.

This paper discusses computational aspects of the distributions of various likelihood ratio test statistics for mean vectors and covariance matrices.

[35] Krishnaiah, P. R. and Yochmowitz, T. G. (1981). Inference on the structure of interaction in two-way classification model. Handbook of Statistics (P. R. Krishnaiah, editor), 973-994. North-Holland Publishing Company.

In this paper, the authors discussed various tests for studying the structure of interactions in two-way classification model with one observation per cell. These test basically involve various functions of the eigenvalues of the Wishart matrix.

[36] Lau, K. and Rao, C. R. (1981). Integrated Cauchy functional equation and characterizations of the exponential law. Sankhya, 44.

A general solution of the functional equation

$$\int_0^{\infty} f(x+y)d\mu(y) = f(x)$$

where  $f$  is a nonnegative function and  $\mu$  is a positive Borel measure on  $[0, \infty]$  is shown to be  $f(x) = p(x) \exp(\lambda x)$  where  $p$  is a periodic function with every  $y \in \underline{\mu}$ , the support of  $\mu$ , as a period. The solution is applied in characterizing Pareto, exponential and geometric distributions by properties of integrated lack of memory, record values, order statistics and conditional expectation.

[37] Lau, K. and Rao, C. R. (1981). Solution to the integrated Cauchy functional equation on the whole line. Sankhya, 44.

A general solution of the integrated Cauchy functional equation

$$\int_{-\infty}^{\infty} f(x+y)d\mu(y) = f(x) \text{ a.e. for } x \in (-\infty, \infty)$$

where  $f$  is a locally integrable positive function and  $\mu$  is a positive Borel measure on  $\mathbb{R}$  is shown to be

$$f(x) = p_1(x)e^{\lambda_1 x} + p_2(x)e^{\lambda_2 x} \text{ a.e.}$$

where  $p_1$  and  $p_2$  are positive periodic functions with every  $u \in \underline{\mu}$ , the support of  $\mu$ , as a period. A variant of the Choquet-Deny theorem on  $\mu$ -harmonic functions is given.

[38] Rao, B. R. and Janardan, K. G. (1981). Characterization theorems involving the generalized Markov-Polya damage model. Technical Report No. 81-04. Institute for Statistics and Applications, Department of Mathematics and Statistics, University of Pittsburgh.

In the present paper, certain random damage models are examined, such as the Generalized Markov-Polya and the Quasi-Binomial, in which an integer-valued random variable  $N$  is reduced of  $N_A$ . If  $N_B$  is the missing part, where  $N = N_A + N_B$ , the covariance between  $N_A$  and  $N_B$  is obtained for some general classes of distributions, such as the G.P.S.D. and M.P.S.D. for the random variable  $N$ . A characterization theorem is proved that under the generalized Markov-Polya damage model, the random variables  $N_A$  and  $N_B$  are independent if, and only if,  $N$  has the Generalized Polya-Eggenberger distribution. This generalizes the corresponding result for the Quasi-Binomial damage model and the generalized Poisson distribution. Finally, some interesting identities are obtained using the independence property and the covariance formulas between the numbers  $N_A$  and  $N_B$ .

[39] Rao, C. R. (1979). Fisher efficiency and estimation of several parameters (Fisher memorial lecture given at the joint IMS-ASA-BS meeting in Washington, DC August 1979).

In this paper, the role of Minimum Mean Square Error as a general criterion for estimation of parameters is critically examined. It is shown that smaller mean square error does not necessarily imply greater concentration of the estimator around the true value. The empirical Bayes method for simultaneous estimation of parameters introduced by Fisher is shown to provide a good ranking of populations for individual populations.

[40] Rao, C. R. and Kleffe, J. (1979). Estimation of variance components. Handbook of Statistics, Vol. 1 1-40. North-Holland Publishing Company.

This paper provides a comprehensive review of the methods of estimation of variance components. Conditions for identifiability and estimability of variance components have been given. Methods of minimum variance unbiased estimation in the normal case, MINQE in the general case and ML estimation in the normal case have been discussed.

[41] Rao, C. R. (1981). Diversity and dissimilarity coefficients: a unified approach. J. Theoretical Population Biology.

Three general methods for obtaining measures of diversity within a population and dissimilarity between populations are discussed. One is based on an intrinsic notion of dissimilarity between individuals and others make use of the concepts of entropy and discrimination. The use of a diversity measure in apportionment of diversity between and within populations is discussed.

[42] Rao, C. R. (1981). Some comments on the minimum mean square error as a criterion of estimation. In Statistics and Related Topics, 123-144. North-Holland Publishing Company.

It is shown that estimators obtained by MMSE (minimizing the mean square error) may not have optimum properties with respect to other criteria such as PN (probability of nearness of the true value in the sense of Pitman) or PC (probability of concentration around the true value). In particular, a detailed study is made of estimators obtained by shrinking the minimum variance unbiased estimators to reduce the MSE. It is suggested that because of mathematical convenience and some intuitive considerations, MMSE could be used as a primitive postulate to derive estimators, but their acceptability should be judged on more intrinsic criteria such as PN and PC.

[43] Rao, C. R. (1961). A lemma on g-inverses of a matrix and computation of correlation coefficients in the singular case. Communications in Statistics A-10, 1-10.

Formulae for multiple, partial and canonical correlations are generally expressed in terms of the elements of the inverse covariance matrix of the variables. These are not valid when the covariance matrix is singular. Appropriate formulae using g-inverses are developed.

The results depend on some basic lemmas on the idempotent matrix  $A^*A$  where  $A^*A = A$  (g-inverse as defined by Rao, 1962) and on the spectral decomposition of a hermitian matrix with respect to a non-negative definite matrix.

[44] Khatri, C. G. and Rao, C. R. (1961). Some generalizations of the Kantorovich inequality. To appear in Sankhya, 44.

Inequalities are obtained for expressions of the type  $|X'AY(Y^{-1}Z)|$ ,  $|X'AY(Y^{-1}AZ)|^{1/2}|XZ|/|Z'AZ|$ ,  $|Z'B^2Z|/|Z'C^2Z|/|Z'BCZ|^2$ , etc., for variations in matrices X and Y given matrices A, B, C when Z and Y are vectors we have the Kantorovich inequality. The inequalities when X and Y are matrices have applications in determining the efficiencies of estimators of several parameters in linear models.

[45] Rao, C. R. (1979). Perspectives in Statistics. Sankhya 41, 129-137. (This is Presidential Address at the meeting of the International Statistical Institute held in Manila.)

The paper refers to "the expanding frontiers of knowledge in statistics and the current controversies, and outlines the role that the International Statistical Institute can play in directing future research to make statistics a socially meaningful and viable science. Suggestions are made about the training of statisticians with

a proper blend of theoretical knowledge and skill in applications, development of statistical courses for specialists in other disciplines, the role of government statisticians, and the use of computers in statistical research. Some examples are given to highlight the difficulties involved in defining the efficiency of an estimator and the possible dangers in the uncritical use of methods developed by academic statisticians in practical work.

[46] Ghosh, J. K. and Subramanyam, R. (1981). Estimation in separated families - two parameter case. Technical Report No. 81-32. Institute for Statistics and Applications, Department of Mathematics and Statistics, University of Pittsburgh.

In an earlier paper Ghosh and Subramanyam (1975) which will be referred as GS (1975) studies some properties of the nle (maximum likelihood estimator) of a discrete parameter. An expression for the asymptotic risk of the nle was obtained in the case of a single discrete parameter. But in most practical cases one has one discrete and one or more continuous parameter. For more examples refer to GS (1975) and Cox (1962). In such a case an expression for the asymptotic risk of the nle of the discrete parameter is given in section 2. Two examples are discussed. The first example is that considered by Cox (1962), deciding between two distributions, Poisson or geometric. Our asymptotic theory is different from that of Cox. The second example is  $\theta(N,p)$ , both  $N$  and  $p$  unknown.

[47] Rao, C. R., Sinha, B. R. and Subramanyam, R. (1981). Third order efficiency of the maximum likelihood estimator in the multinomial distribution. Technical Report No. 81-21. Institute for Statistics and Applications, Department of Mathematics and Statistics, University of Pittsburgh.

Consider a multinomial population with  $k$  cells. Let  $\pi_1(0), \dots, \pi_k(0)$  be the population proportions and  $p_1, \dots, p_k$  be the sample proportions out of a

sample of size  $n$ . We restrict our attention to the class of estimators  $T_n$  obtained as the solution of the equation:

$$\hat{\theta}_n \in \left\{ \frac{p}{T_n, T_n^2} \right\} \hat{\theta}_n'(T_n).$$

We show that a.s.e. condition for  $\hat{\theta}_n$  to be second order efficient (SSE) is that  $\hat{\theta}'(t) \neq 0$ ,  $\hat{\theta}''(t) = 0$ . In the subclass of SSE estimators let  $T_n$  by any estimator such that the corresponding  $\hat{\theta}'''(t) \neq 0$ . It is proved that there exist bias-adjusted versions of  $\hat{\theta}_n$ , the ase, and  $T_n$ , say  $\hat{\theta}_n^{(1)}$  and  $T_n^{(1)}$ , such that  $\hat{\theta}_n^{(1)}$  has a smaller mean squared error up to  $O(n^{-2})$  than  $T_n^{(1)}$ . Here  $\hat{\theta}_n^{(1)}$  and  $T_n^{(1)}$  have the same bias up to  $O(n^{-2})$ . This generalizes the SSE property of the ase proved earlier by Rao (1963).

[48] Rao, C. R. (1981). Gini-Simpson Index of diversity: a characterization, generalization and applications. Biometrika, 68, 11.

The Gini-Simpson Index of diversity of a multinomial distribution defined by k cell probabilities  $p = (p_1, \dots, p_k)$  is  $I = 1 - p_1^2 - \dots - p_k^2$ . In this paper, this index is characterized by a set of postulates based on a measure of distance between distributions. Further, it is suggested that in practical applications involving measurement of qualitative variation it is more appropriate to use an Index of the type  $p' \Delta p$  where  $\Delta$  is a  $k \times k$  matrix of generalized distances between attributes characterizing the cells of the multinomial distribution. The Gini-Simpson Index corresponds to a special choice of  $\Delta$ . The expression  $p' \Delta p$  may also be interpreted as a generalized quadratic entropy. An example is worked out illustrating analysis of diversity and cluster analysis based on the measure  $p' \Delta p$ .

[49] Rao, C. R. (1981). Analysis of diversity: a unified approach. Technical Report No. 81-20, Institute for Statistics and Applications, Department of Mathematics and Statistics, University of Pittsburgh.

A general method of constructing a measure of diversity within a population subject to some postulates is described. It is shown that such a diversity measure, when it satisfies a concavity condition, can be used to analyse diversity in a mixture of populations as due to main effects and interactions of factors, or the levels of which the individual populations are specified. The method is applicable to both quantitative and categorical response data, and provides a generalization of the technique of analysis of variance. Some applications are considered.

[50] Rao, C. R. (1981). Optimization of functions of matrices with applications to statistical problems. Technical Report No. 81-31, Institute for Statistics and Applications, Department of Mathematics and Statistics, University of Pittsburgh.

The paper reviews some methods of optimizing functions of vectors and matrices subject to some restrictions and develops techniques for solving them without using the calculus of matrix derivatives. Their application to a number of statistical problems in linear estimation and multivariate analysis is illustrated.

[51] Rao, C. R. (1982). Diversity: Its measurement, decomposition, apportionment and analysis. *Sankhyā*, 44 (in press).

Two general methods of obtaining measures of diversity within a population are discussed. One is based on an intrinsic notion of dissimilarity between individuals and the other makes use of the concepts of entropy. Some examples are given of the decomposition of diversity within a population in terms of given or conceptual

factors.

Methods for computations of diversity in a hierarchically classified set of populations are discussed.

The concept of analysis of diversity as a generalization of analysis of variance is developed for populations classified by combinations of different levels of chosen factors.

(iii) Sampson, A. (1980). Nonnegative Cholesky decomposition and its application to association of random variables. *Statist. J. of Mathematics and Statistics Methods* 1: 785-791.

The concept of a multivariate family of distributions indexed by a covariance scale parameter  $\lambda$  is formally defined and examples are given. The multivariate analog to one such family, sufficient conditions are given so that a positive definite matrix has a nonnegative Cholesky decomposition. These conditions also yield the association of random variables with a covariance scale parameter distribution. These results are related to other matrix results and to Barlow and Proschan's stronger conditions (Statistical Theory of Reliability and Life Testing: Probability Models, Holt, Rinehart, Winston, New York, 1973), for the association of the multivariate analog, namely,  $t_{ij} \geq 0$ ,  $i \neq j$  where  $t^{(1)} = \lambda + t_{11}$ .

(iv) Sampson, A. A. (1980). Positive dependence properties of elliptically symmetric distributions. Accepted subject to revision to *Journal of Multivariate Analysis*.

Let  $x_1, \dots, x_p$  have p.d.f.  $g(x_1^2, \dots, x_p^2)$ . It is shown that (a)  $x_1, \dots, x_p$  are positively lower orthant dependent or positively upper orthant dependent if, and only if,  $x_1, \dots, x_p$  are i.i.d.  $\mathcal{N}(0, \sigma^2)$ ; and (b) the p.d.f. of  $|x_1|, \dots, |x_p|$  is pp if, and only if,  $g(u)$  is convex. Let  $x_1, x_2$  have p.d.f.  $f(x_1, x_2) = g(x_1^2, x_2^2) t^{-1}(x_1^2, x_2^2)^{\alpha}$ . Necessary and sufficient conditions are given for  $f(x_1, x_2)$  to be pp for fixed correlation  $\rho$ . It is shown that if  $t$  is pp for all  $\alpha > 0$ , then  $(x_1, x_2)^T \sim \mathcal{N}(0, I)$ . Related positive dependence

results and applications are also considered.

(14) Sampson, A. R. (1988). A Multivariate  $\alpha$  Ratio. Technical Report No. 88-6, Institute for Science and Applications, Department of Mathematics and Statistics, University of Pittsburgh.

A brief review of the historical background and certain basic results concerning the univariate correlation ratio are given. A multivariate correlation ratio of a random vector  $T$  upon a random vector  $S$  is defined by  $\alpha_{\Lambda}(T,S) = \text{trace}^{-1}(\text{cov}(ST)) / \text{trace}^{-1}(S^2)$ , where  $\Lambda$  is a given positive definite matrix. The properties of  $\alpha_{\Lambda}$  are discussed, with particular attention paid to a "correlation-correlating" property. A number of examples include the multivariate normal, the elliptically symmetric distributions, and the multinormal. The problem of calculating  $\alpha_{\Lambda}(ST,S)$  over suitable matrices  $\Lambda$  is considered and the results that are obtained are related to canonical correlations for the multivariate normal.

(15) Sampson, A. and Smith, R. (1988). Assessing risks through the determination of rare event probabilities. Accepted subject to revision in Geological Research.

We consider the problem in risk assessment of evaluating the probability of occurrence of rare, but potentially catastrophic, events. The lack of historical data due to the sheer novelty of the event makes conventional statistical approaches inappropriate. The problem is compounded by the complex multivariate dependencies that may exist across potential event sites. In order to evaluate the likelihood of one or more such catastrophic events occurring, we provide an information theoretic model for merging a decision maker's opinion with expert judgment. Also provided is a methodology for the reconciling of conflicting expert judgments. This merging approach is invariant to the decision maker's viewpoint in the limiting case of exceptionally rare events. These methods are applied to certain case studies.

[134] Thompson, A. B. (1981). Stochastic approximation. To appear in Encyclopedie of Statistical Sciences. John Wiley and Sons.

This article reviews univariate and multivariate stochastic approximation procedures from the Robbins-Robbins and Stochastic-Halton procedures to nearly definite nonstochastic estimation procedures. Variations of these procedures are also described. Applications of stochastic approximation procedures are also considered.

[135] Stukel, D. R. and Raghavarao, D. (1981). Estimation of the common mean of two univariate normal populations. Technical Report No. 81-07, Institute for Mathematics and Applications, Department of Mathematics and Statistics, University of Pittsburgh.

The problem of estimating the common mean  $\mu$  of two univariate normal populations with unknown and unequal variances is considered from a decision-theoretic point of view. We restrict our attention to an appropriate class  $C$  and its three subclasses  $C_0$ ,  $C_1$ ,  $C_2$  of admissible estimates of  $\mu$ . It is shown that the usual estimate of  $\mu$ , which is the weighted linear combination of the sample means with weights as reciprocals of the sample variances, is admissible in  $C_0$  and extended admissible in  $C$  but is neither Bayes nor limiting Bayes in  $C_1$ . Admissible estimates in  $C_1$  and  $C_2$  are obtained. The loss is always assumed to be squared error.

[136] Stukel, D. R. and Sather, S. E. (1982). Invariant confidence sequences for some parameters in a general linear regression model.

Let  $\mathbf{z}_1, \mathbf{z}_2, \dots$  be independent  $p$ -variate normal vectors with  $E \mathbf{z}_n = \mathbf{0} \mathbf{I}_p$ ,  $n = 1, 2, \dots$  and some p.d. dispersion matrix  $\Sigma$ . Let  $\alpha$ ,  $p$ ,  $q$  and  $t$  be unknown parameters and  $\mathbf{y}_n$ 's are known  $q+1$  vectors. Writing  $\boldsymbol{\theta} = (\theta_1' \ \theta_2')' = (\theta_{(1)}' \ \theta_{(2)})'$  with  $\theta_1' = p_1^{-1}(p_1^{-1}q_2^{-1}p)$  and  $\theta_{(1)}' = p_1^{-1}q_1^{-1}(q_1^{-1}q_2^{-1}q)$ , we have constructed invariant confidence sequences for (i)  $\theta$ , (ii)  $\theta_{(1)}$ , (iii)  $\theta_2$ , when  $\theta_2 = 0$  and (iv)  $\theta^2 = \{t\}$ . This uses the basic ideas of Robbins (1970) and generalizes some of his and Lei's (1976) results. In the process

alternative simpler solutions of some of Khan's results  
(1978) are obtained.